DOI：10.13998/j.cnki.issn1002-1248.23-0312

地理标志农产品品牌管理政策注意力分析模型构建及应用

霍梦佳[1]，刘　娟[1,2*]，黄　杰[1]

（1.中国农业科学院农业信息研究所，北京　100080；2.农业农村部 农业大数据重点实验室，北京　100080）

摘　要：[目的 / 意义]地理标志是中国地方政府开展品牌建设的重要抓手，政府在地理标志农产品品牌管理领域的政策注意力问题值得深入探究。论文旨在构建地理标志品牌分析模型，以支持分析品牌发展现状、探究政策注意力分布、提出优化策略。[方法 / 过程]论文首先采集数据构建语料库，采用 UIE 通用信息抽取框架抽取文本，采用基于 Transformer 模型的品牌管理措施分类模型，开展管理措施分类，最终完成构建品牌分析模型，支持政策注意力分析，找到瓶颈问题，进行措施优化等应用。[结果 / 结论]论文以烟台苹果为例进行模型验证，对大量语料进行抽取、分类，发现烟台苹果政策注意力较集中、措施雷同性高，收集文本中的 41.1%集中在品牌定位与规划部分，31.7%集中在品牌核心能力提升部分，涉及品牌营销及保护的不到 10%，未利用好地理标志品牌的品牌效应。

关键词：农产品地理标志；自然语言处理；政策注意力；信息抽取；文本分类

中图分类号：F303.3　　　文献标识码：A　　　文章编号：1002-1248（2023）07-0094-11

引用本文：霍梦佳，刘娟，黄杰. 地理标志农产品品牌管理政策注意力分析模型构建及应用[J]. 农业图书情报学报，2023，35（7）：94-104.

1 引　言

2022 年农业农村部办公厅关于印发《农业品牌精品培育计划（2022—2025 年）》（农市发〔2022〕8 号），启动 2022 年农业品牌精品培育工作，提出加快促进区域农业品牌在品牌基础、营销推广、管理服务、市场消费等方面全面提升[1]。农产品地理标志是农业品牌重要组成部分，标识农产品来源于特定地域，用以传递产品质量等属性方面的“信号”，灵活应用原产地效应，对国家和农产品原产地的经济发展而言，有着积极的意义[2]。自中国农业部 2008 年正式启动农产品地理标志等级保护工作起，至 2022 年 4 月底中国已登记农产品地理标志品牌 3 510 个，地理标志农产品品牌注

收稿日期：2023-05-18

基金项目：中国农业科学院基本科研业务费项目“基于文本挖掘的地理标志农产品品牌管理模式识别技术研究与应用”（JBYW-AII-2022-33；Y2023XK13）；现代农业产业技术体系北京市创新团队建设项目“现代农业产业技术体系北京市创新团队建设”（BAIC10-2023-E10）

作者简介：霍梦佳（1997- ），女，研究生，研究方向为农业数据分析挖掘。黄杰（1997- ），男，博士研究生，研究方向为自然语言处理

* 通信作者：刘娟（1978- ），女，博士，副研究员，研究方向为农业数据挖掘。Email：liujuan@caas.cn

册登记工作取得了巨大成就，但中国地理标志农产品品牌建设经营还比较弱。2022 年中国农业品牌研究中心发布首个农业品牌行业标准《农产品区域公用品牌建设指南》（下称《指南》）[3]，提出农产品区域公用品牌建设五大关键步骤。各地政府重视品牌建设但政策如何高效发力问题有待深入探讨。基于品牌注意力研究分析区域农产品品牌管理中政策偏向性、全面性，确定品牌优化方向，对于各地政府优化农产品地理标志品牌管理至关重要。

2　文献综述

学术界关于地理标志农产品的研究主要是依据理论研究、空间计量分析、立法研究、案例探究等方法，从品牌发展、品牌保护、品牌困境等角度入手，探究制约农产品地理标志发展的因素，并提出相应的对策，但缺少从大数据角度研究，对现有政策资源的整合、分析、利用不足。随着互联网海量数据的增加，在政策注意力研究领域的大数据分析应用逐渐增多，但大多分为以下 3 类：一是基于现有的文本分析工具对主题网络进行分析[4,5]，只能对热点词进行分析，不能体现热点词间的关系；二是基于扎根理论对文本进行编码进而分析政策注意力[6,7]，然而对大量文本进行编码耗时耗力；三是基于内容分析法对具体的政策内容进行分析[8,9]，单个的政策研究不能展示所有政策范围及注意力分布。此外，当前研究也面临在海量的互联网信息中收集高度相关的文本构建语料库，以及训练适用于地理标农产品品牌领域的文本研究模型的新挑战。

因此，本研究将领域知识与现有的文本挖掘技术相结合，融合领域知识构建了一套基于文本挖掘技术的地理标志农产品品牌管理政策注意力分析模型；并以烟台苹果为例，从定量角度分析该产业目前的政策注意力分布，发现品牌发展瓶颈，并提出相对应的优化策略。

3　研究方法

本研究主要分为两个处理过程：信息抽取部分和注意力分析部分。

其中信息抽取部分为命名实体识别和文本抽取两部分，其中命名实体识别部分采用基于字典匹配的方式抽取文本中包含的地标品牌，如“烟台苹果”；信息抽取部分抽取文本语料中包含的管理措施类信息，基于以上两部分信息抽取结果，构建农产品地理标志品牌管理措施库。

3.1　命名实体识别技术

由于本研究具有完整的农产品地理标志名称目录，能够构建完整有限的词典，且该类实体没有简写、组合、歧义等特殊情况，所以采用基于词典的命名实体识别方法能够获得较高的准确率，提升识别效率。基于词典的实体识别方法是指词典中的每个词与被处理文档之间逐一匹配的过程，在不需要发现未知词的命名实体识别任务中，基于词典的匹配方式具有较高的优越性。

3.2　文本抽取技术

本研究的抽取目标是抽取品牌的相关管理措施，如“依托灌云杨集豆丹养殖协会建设 600 亩核心养殖基地”等。由于措施文本长短不一，为优化抽取效果，采用更加灵活的基于中文通用信息抽取框架的方式。中文通用信息抽取框架 UIE 是基于 ERNIE 3.0 增强预训练模型训练的，支持下游任务训练微调（图 1）。

（1）ERINE 3.0 模型。ERNIE 3.0[10]基于 Transformer-XL[11]设计了一种新的持续多范式统一的预训练框架，解决了普通 Transformer 无法建立超过固定长度文本的长依赖和上下文碎片化问题，使预测的效率也大幅提升。经验证，模型在 20 多个不同类型的自然语言处理任务上取得了 3%以上的显著提升。

（2）通用信息抽取框架（Universal Information Extraction，UIE）。基于所有的信息抽取任务都可以建模为在文本中寻找目标信息片段和确定目标信息片段之间的关系两个子过程[12,13]，UIE 将所有信息抽取任务通过结构化模式提示器（Structural Schema Instructor，SSI）生成特定抽取任务的 Schema，再通过结构化抽取语言（Structured Extraction Language，SEL）把不同

DOI：10.13998/j.cnki.issn1002-1248.23-0312

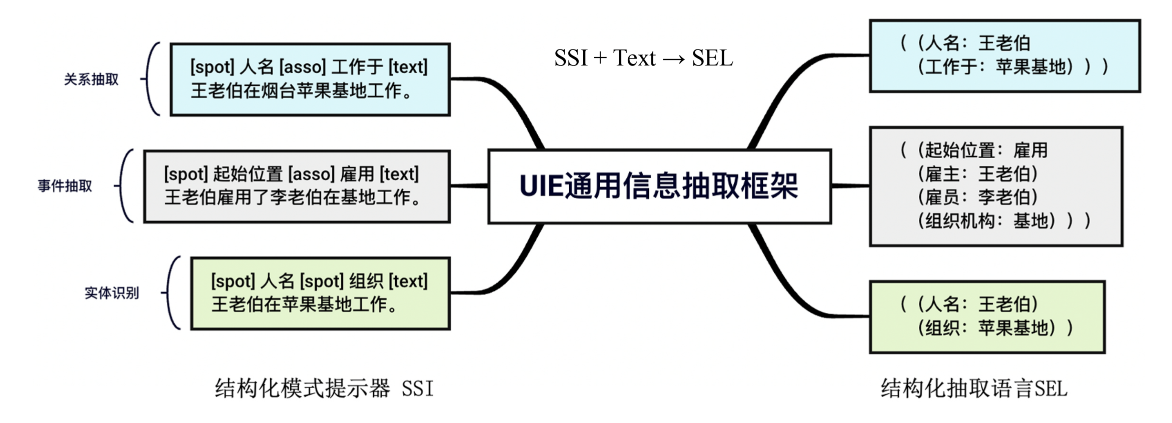


图 1　UIE 通用信息抽取框架示意图

Fig.1　Diagram of UIE general information extraction framework

任务的抽取结果统一用一种语言表示。

基于 ERNIE 3.0 知识增强预训练模型训练的中文通用信息抽取模型 UIE 可以支持不限定行业领域和抽取目标的关键信息抽取，并提供基于小样本微调功能，大幅降低标注数据的依赖，降低成本，提升效果。

3.3　文本分类技术

政策注意力分析部分的主要目标是将已抽取出来的措施文本分类，得到政府的政策注意力分布。本研究采用基于 Transformer 的文本多分类模型，Transformer[14]完全采用自注意力机制，利用自注意力机制实现了并行计算，大大提高了模型的运行效率。Transformer 本质是一个 Encoder-Decoder 结构，编码端和解码端均由 6 层组成，Transformer 的输入为单词的 Embedding 和单词位置的 Embedding 相加的结果，经过 6 个 Encoder 层后得到句子的所有单词的编码矩阵；再将编码矩阵信息输入进 Decoder 中，经过处理后输出结果。由于 Transformer 本身不利用单词的顺序信息，因此需要再输入中加入位置信息，加之自注意力机制，可以捕获单词之间多种维度上的相关系数[15]。

4　品牌管理政策注意力分析模型构建

本研究基于文本挖掘技术并结合领域知识构建了一套地理标志农产品品牌管理措施研究模型，模型最初始输入为品牌政策文本，通过两个模块获得最终输出为政策注意力分析结果图，两个模块主要流程如下：一是信息抽取模块，输入为已构建的语料库，此模块输出的结果为各地理标志农产品品牌的管理措施库；二是文本分类模块，此部分将抽取出的结果作为输入，输出政策注意力分布图，模型结构如图 2 所示。

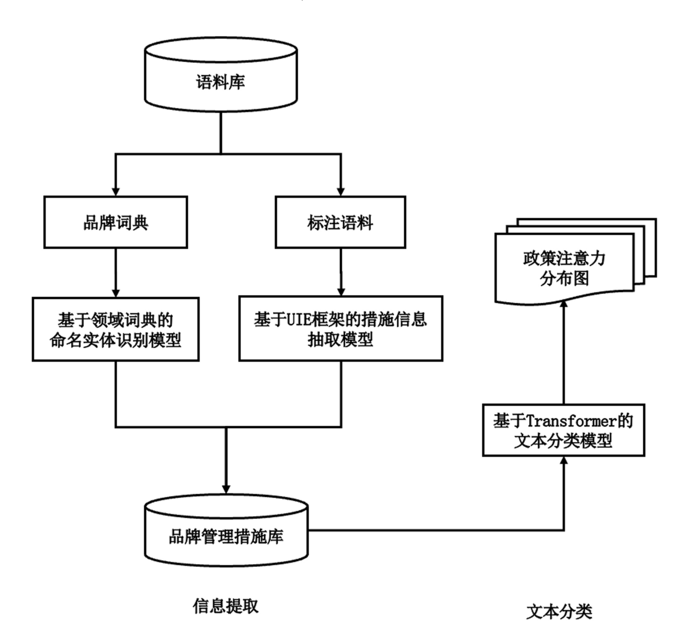


图 2　地理标志农产品品牌管理政策注意力分析模型框架

Fig.2　Attention analysis modeling framework for the brand management policy of agricul-tural products with geographical indications

4.1　数据集

4.1.1　数据源选择

互联网信息资源丰富，经过调研多个农业类网站

后发现大部分农业网站包含相关内容较少、分布较散且网页结构化程度较低，若全文爬取将会出现极大的噪音数据，这将极大的增加筛选工作的工作量，并影响后续的文本抽取准确率，因此本研究将爬取源定为少部分专业类网站及各省的政府门户网站，以提升语料采集质量。

（1）全国农产品地理标志查询系统。全国农产品地理标志查询系统（http://www.anluyun.com/）是在农业农村部积极推进地理标志农产品登记、追溯、管理的背景下建设的地理标志农产品追溯平台。平台以“互联网＋农业”为手段基于物联网、云计算、地理信息 GIS 等现代信息技术使系统内部信息能够达到全面准确、更新及时，可以实现追溯数据共享，互联互通，全面提升了农产品质量追溯信息化管理工作水平。本研究为保证品牌信息涵盖范围的全面性、准确性、零滞后性，选定全国农产品地理标志品牌查询系统作为品牌基本信息的来源。

（2）中国农业品牌公共服务平台。中国农业品牌公共服务平台（http://aboc.agri.cn/）由农业农村部信息中心主办，平台以农产品区域公用品牌、企业品牌、产品品牌为核心，围绕供应链、产业链、数据链，提供丰富详实、权威及时、实用有效的信息。平台设立“品牌资讯”模块，提供各地农产品品牌管理发展的相关新闻报道，数据真实权威、更新及时、覆盖面广，符合农产品地理标志品牌管理措施语料库的构建标准。

（3）各省的政府门户网站。各省政府网站是了解各地政府政务的最直接平台。经过调研各地政府网均包含地理标志相关报道，且内容权威有效、实时性强，并且各地政府网站提供信息检索功能、支持互联网文本爬取技术，因此本研究在各地政府网站中以“农产品地理标志”为检索词的检索结果作为品牌管理措施库的来源。

4.1.2　数据集构建

基于以上网上数据爬取结果共构建两个数据库。

（1）农产品地理标志品牌库。此步共搜集 3 510 个品牌名称，支持后续构建领域词典，数据展示如图 3 所示。

（2）农产品地理标志措施库。基于以上数据源，利用 Python 语言中获取网络资源的框架 Scrapy 获取结

产品名称	产地	产品编号	证书持有者	登记年份
黎城核桃	山西省长治市	AGI00001	黎城县核桃产业办公室	2008 年
芮城花椒	山西省运城市	AGI00029	山西省芮城县花椒产业协会	2008 年
红山荞麦	山西省朔州市	AGI00030	朔州市平鲁区红山荞麦协会	2008 年
长子大青椒	山西省长治市	AGI00031	山西省长子县椒王蔬菜营销合作社	2008 年
孝义核桃	山西省吕梁市	AGI00032	山西省孝义市农业技术推广中心	2008 年
天山大明绿豆	内蒙古自治区赤峰市	AGI00033	内蒙古赤峰市阿鲁科尔沁旗农业环保能源工作站	2008 年
乌兰察布马铃薯	内蒙古自治区乌兰察布市	AGI00034	乌兰察布市农畜产品质量安全中心	2008 年
交城骏枣	山西省吕梁市	AGI00002	交城县农业技术推广中心	2008 年
乌珠穆沁羊肉	内蒙古自治区锡林郭勒盟	AGI00035	锡林郭勒盟农牧业科学研究所	2008 年
乌海葡萄	内蒙古自治区乌海市	AGI00003	乌海市植保植检站	2008 年
丹东杜鹃	辽宁省丹东市	AGI00036	丹东市花卉协会	2008 年
小梁山西瓜	辽宁省沈阳市	AGI00037	新民市梁山镇西瓜协会	2008 年
彰武黑豆	辽宁省阜新市	AGI00004	彰武县农业发展服务中心	2008 年
集安五味子	吉林省通化市	AGI00038	吉林省集安北五味子产业协会	2008 年
黑山锦丰梨	辽宁省锦州市	AGI00005	黑山县农业技术推广中心	2008 年
洮南辣椒	吉林省白城市	AGI00039	洮南市辣椒产业协会	2008 年
巴彦猪肉	黑龙江省哈尔滨市	AGI00040	巴彦县农产品质量安全协会	2008 年
岫岩滑子蘑	辽宁省鞍山市	AGI00006	岫岩满族自治县蔬菜生产管理局	2008 年
阿城大蒜	黑龙江省哈尔滨市	AGI00041	哈尔滨市阿城区金源绿色农畜产品协会	2008 年
呼兰大葱	黑龙江省哈尔滨市	AGI00042	哈尔滨市呼兰区兰河街道办事处社区服务中心	2008 年
嘉荫大豆	黑龙江省伊春市	AGI00043	嘉荫县绿色食品协会	2008 年
洮南绿豆	吉林省白城市	AGI00007	洮南市杂粮杂豆经销协会	2008 年
峄城石榴	山东省枣庄市	AGI00044	峄城区标准化农业产业协会	2008 年
新开河贡米	吉林省通化市	AGI00008	集安绿色食品产业协会	2008 年
昌乐西瓜	山东省潍坊市	AGI00045	昌乐县农产品质量检测中心	2008 年
集安边条参	吉林省通化市	AGI00009	集安绿色食品产业协会	2008 年

图 3　农产品地理标志品牌库

Fig.3 Basic information database of agricultural products with geographical indications

DOI：10.13998/j.cnki.issn1002-1248.23-0312

果，主要使用 Scrapy 框架的爬虫（Spider）组件设置爬取策略。Spider 是 Scrapy 用户编写用于分析 Response 并提取 Item 或额外跟进的 URL 的类，具有高效快捷获取网络资源的特点。

经过初步的去掉空文本和重复文本的预处理过程，最终获取共包括相关文本共 8 415 条语料数据，数据库如图 4 所示。

4.2 文本抽取模型构建

本研究将非结构化文本的抽取分为命名实体识别和信息抽取两部分，其中命名实体识别部分抽取文本中包含的品牌名称，如“烟台苹果”等；文本信息抽取部分抽取文本语料中包含的管理措施类信息，如“依托灌云杨集豆丹养殖协会建设 600 亩核心养殖基地”“组织参加农博会、绿展会、农洽会等国内外大型农产品展销活动”等，抽取框架如图 5 所示，包含两个部分：命名实体识别和信息抽取，两部分的输入均为文本语料，分别输出文本包含的实体名称及措施。

4.2.1 命名实体识别模型

地理标志农产品品牌领域的文本与普通领域的文本有些区别，主要体现在该领域有完整的品牌名称词典，因此基于领域词典来构建命名实体识别模型会大

title	content
王大嫂为地理标志农产品城口山地鸡代言	11月28日上午，第六届全国农产品地理标志品牌推介会在悦来国际会议中心举行。本届农产品地理标志专
“中国农产品地理标志品牌形象馆”对外发布	近日，“中国农产品地理标志品牌形象馆”正式上线面向全球发布。地理标志农产品是世界通行的文化语言，
“开封县花生”地理标志农产品亮相2021年丰收节河南主会场	9月23日，以“庆丰收、感党恩”为主题的2021中国农民丰收节河南省主会场在焦作武陟嘉应观开幕。祥
陈疃镇：加快打造“日照蓝莓”地理标志品牌	编者按惟创新者进，惟创新者强，惟创新者胜。今年，山东把创新摆在发展全局的核心位置，重点抓好“
25家咖企获“普洱咖啡”地理标志证明商标使用授权	近日，普洱咖啡协会为首批获授权使用“普洱咖啡”地理标志证明商标的咖啡企业颁发地理标志产品授权书
江苏灌云县实施豆丹地理标志保护工程提升产品竞争力	一年四季皆有豆丹上市，年产值达30亿元的灌云豆丹产业，通过实施“灌云豆丹”地理标志农产品保护工
“新化红薯粉”成功注册地理标志证明商标	日前，国家知识产权局发布公告，正式核准“新化红薯粉”注册地理标志证明商标。这是娄底继2021年获
湖北随州17家企业获准使用国家地理标志	近日，湖北省知识产权局发布公告，核准随县萌阳食品有限公司等5家企业（合作社）使用国家地理标志
广西柳州螺蛳粉请来专家助力 加快推进地理标志品牌化建设	中国质量新闻网讯 为加快建立柳州螺蛳粉地理标志品牌建设标准体系，推动柳州螺蛳粉产业高质量发展，
“灵山奶水牛”入选地理标志登记产品	近日，农业农村部公布2022年第一批农产品地理标志登记产品公告信息，灵山县“灵山奶水牛”榜上有名。
安达打造强农“地理标志”品牌	日前，“安达黑豆”由国家知识产权局核准注册成功，安达地理标志标识商标增加到4件，其中地理标志证
“宁夏枸杞”地理标志证明商标正式授权使用	新华社客户端宁夏频道4月15日电（记者赵倩）日前，“宁夏枸杞”地理标志证明商标授标启动仪式在银川
“金溪蜜橘”获批国家地理标志证明商标	4月2日，金溪县市场监督管理局将“金溪蜜橘”地理标志证明商标颁发给金溪县蜜橘产业协会，标志着“金
新疆新增12家可使用国家地理标志专用标志企业	近日，经国家知识产权局核准，新疆生产哈密瓜、吐鲁番葡萄干、和田大枣等6种产品的12家企业可使用
“防城肉桂”获评国家地理标志证明商标	近日，国家知识产权局发布公告，对防城港市防城区申请注册“肉桂（调味品）”的地理标志证明商标“防
“鄂托克前旗羊肉”品牌 入选国家地理标志运用促进重点名录	近日，记者从内蒙古自治区鄂尔多斯市鄂托克前旗人民政府获悉，“鄂托克前旗羊肉”成功选入国家地理标
江东时评 | 用好“地理标志”打响“特色品牌”	江苏有“鱼米之乡”的美誉，背靠自然禀赋，孕育出了不少地标产品。近日，淮安市“蒋坝螺蛳”“涟水草鸡
农产品冠上地理标志，“国字号” 特色文化助力乡村振兴	农产品冠上地理标志“国字号” 特色文化助力乡村振兴来源 | 农业农村部近日，中欧地理标志协定生效，南
如皋“新官紫桃”获批国家农产品地理标志登记产品	近日，农业农村部发布公告，“新官紫桃”获批国家农产品地理标志登记。这是如皋市继“如皋黄鸡”“下原
“宜陵螺蛳”“高邮湖白虾”获地理标志商标	扬州网讯（通讯员 杨宇 柏玮娜 记者 邱凌） 近日，扬州市江都区宜陵农村供销合作经济联合会和高邮湖大
黑龙江省农产品地理标志巡礼九：虎林椴树蜜	虎林椴树蜜——完达山脉野生蜜源，纯天然绿色食品。虎林市蜂业协会拥有的“虎林椴树蜜”品牌，成功入
黑龙江省农产品地理标志巡礼十一：阿城粘玉米	近年来，“阿城粘玉米”获得了可喜的成果。2017年获得消费者最喜爱的中国农产品区域公用品牌、黑龙江
潍坊6件地理标志商标入选山东省重点地理标志保护清单	潍坊日报社潍坊融媒讯近日，省市场监管局印发全省重点地理标志保护清单，潍坊市桂河芹菜、诸城绿茶、
绥化市地理标志商标增加至16件 数量位居黑龙江省第一	黑龙江网讯（记者 刘雨珊）“安达黑豆”地理标志证明商标经绥化市、安达市两级市场监督管理局持续培育
枣庄市5件地理标志产品入选省重点地理标志保护清单	近日，山东省市场监管局印发《关于确定山东省重点地理标志保护清单的通知》，其中山亭地瓜枣、山亭

图 4　农产品地理标志品牌管理措施语料库

Fig.4 Corpus of brand management measures of agricultural products with geographical indications

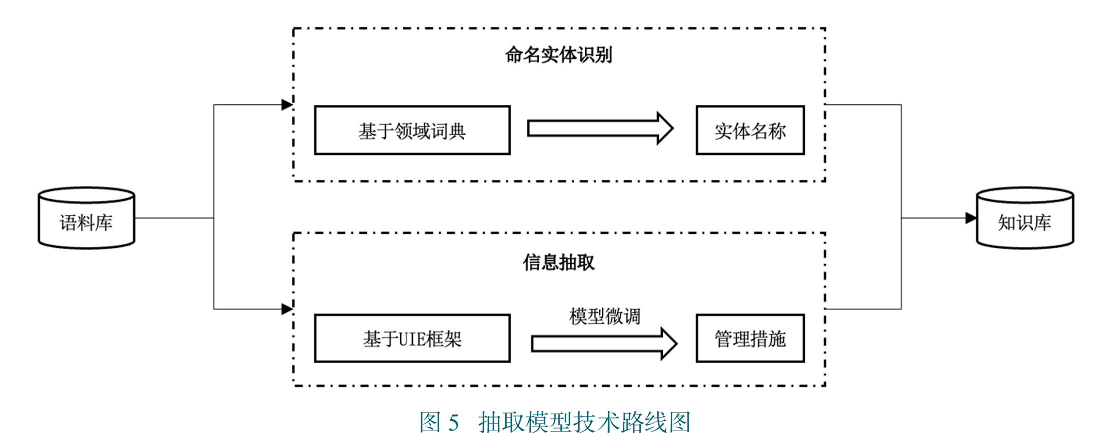


图 5　抽取模型技术路线图

Fig.5 Technology roadmap of the extraction model

大提高命名实体识别准确率。领域词典来源主要是全国农产品地理标志查询系统，通过爬虫技术获得完整的地理标志品牌名录，共包括 3 510 个名称类实体，词典如表 1 所示。

4.2.2 措施文本信息抽取模型

（1）语料标注。语料标注是文本信息抽取的基础性工作，只有高质量的标注语料才能更好地训练出符合要求的模型。在以往的文本信息抽取工作中，语料标注是极耗费人力、时间的一项工作，然而由于本研究使用的模型支持小样本训练，只需要标注少数语料即可到达较好的抽取效果。本次抽取任务的目标为抽取各地具体的管理措施，故本次抽取任务仅需抽取“action”这一个标签，代表措施。我们仅需对文本语料中的措施内容进行标注，语料标注样例如下，在样例段落中，“建立健全地理标志产品保护机制和监督管理办法”“完善地理标志产品技术标准、检验检测和质量保证‘三大体系’”等均为具体的措施，由于计算机识别长句子的效果较差，故本次标注按逗号切割。最后将语料标注为以下格式，包括语料 id、文本内容 text、标签 label、起始位置、终止位置，标注语料文件最后以 json 格式存储（图 6）。

（2）实验设置。实验前需对输入数据进行数据转换，转换成模型适用的格式，并将数据集按 8:1:1 划分为训练集、验证集和测试集。数据准备工作完成后，将实验训练参数设置如表 2 所示。

表 2 实验训练参数

Table 2 Experimental training parameters

参数名称	参数值
learning_rate	1e-5
batch_size	16
max_seq_len	512
num_epochs	100
seed	1 000
valid_steps	100
Optimizer	Adam
device	GPU

其中，learning_rate 代表学习率，默认为 1e-5；batch_size 代表批处理大小，设置为 16；max_seq_len 代表文本最大切分长度，输入超过最大长度时会对输入文本进行自动切分，设置为 512；num_epochs 代表训练轮数，设置为 100；seed 代表随机种子，设置为 1 000；valid_steps 代表评估的间隔 steps 数，设置 100；Optimizer 代表实验选用的优化器，默认为 Adam；device 代表选用训练的设备，设置为 GPU 设备。

（3）模型评估。参数的选取会对抽取的结果产生一些影响，为了使训练的模型获得最佳的抽取效果，本研究基于已构建语料选取不同的实验参数进行比较

表 1 农产品地理标志品牌词典

Table 1 Dictionary of geographical indication brands for agricultural products

品牌名称	品牌名称	品牌名称	品牌名称
黎城核桃	天山大明绿豆	丹东杜鹃	洮南辣椒
芮城花椒	乌兰察布马铃薯	小梁山西瓜	巴彦猪肉
红山荞麦	交城骏枣	彰武黑豆	岫岩滑子蘑
长子大青椒	乌珠穆沁羊肉	集安五味子	阿城大蒜
鄂尔多斯细毛羊	莎车巴旦姆	安和香芋	……

"id":1,
"text":"建立健全地理标志产品保护机制和监督管理办法，完善地理标志产品技术标准、检验检测和质量保证“三大体系”；开展日常监管和年度专项监督检查，严厉查处地理标志违法行为，保护社会公共品牌，维护地理标志产品形象和声誉。",
"label":[[0,21,"action"],[22,51,"action"],[52,67,"action"],[68,80,"action"],[81,89,"action"],[90,103,"action"]]

图 6 语料标注样例

Fig.6 Sample of corpus annotation

DOI：10.13998/j.cnki.issn1002-1248.23-0312

训练。学习率对于深度学习来说尤为重要，一个合适的学习率不仅能加速训练的拟合，还能更好地逼近最优解，本研究针对不同的学习率进行对比实验。

在其他参数均为默认值的前提下，设置不同的学习率进行实验，在学习率为 1e-6 到 5e-4 的范围内进行训练，得到结果如图 7 所示，横坐标为学习率的值，纵坐标为不同学习率下各项实验指标的结果。

训练结果表明，当学习率在 1e-6 到 5e-4 的范围内进行训练时，准确率、召回率和 *F*1 的值显示波动的大致趋势。当学习率在 1e-5 时，*F*1 值达到最优。所以选取学习率为效果较优的 1e-5 进行后续任务抽取。

4.2.3　文本抽取结果展示

利用已构建的命名实体识别模型和信息抽取模型，基于已确定的实验参数对语料库文本进行抽取。将文本按条输入已构建的模型中，因为文本均为长篇语料，故按句分割，按句输出抽取结果，输入案例如图 8 所示。

一年四季皆有豆丹上市，年产值达 30 亿元的灌云豆丹产业，跑出高质量发展“加速度”。今年以来，灌云县通过实施“灌云豆丹”地理标志农产品保护工程，提升了生产能力、标准化程度和品牌效应，建立销售全产业链条的“灌云豆丹”现代化示范基地。

图 8　文本示例

Fig.8　Text sample

将上述文本输入进我们已构建的模型，通过命名实体识别和措施抽取两个流程，获得目标内容，结果框架包含品牌名称及措施，示例结果展示如图 9 所示。代表“灌云豆丹”这一地理标志品牌的管理措施包括“实施‘灌云豆丹’地理标志农产品保护工程”“建立销售全产业链条的‘灌云豆丹’现代化示范基地”等。

{
"名称": "灌云豆丹",
"措施":
[
"实施“灌云豆丹”地理标志农产品保护工程",
"提升了生产能力、标准化程度和品牌效应",
"建立销售全产业链条的“灌云豆丹”现代化示范基地"
]

图 9　抽取结果示例

Fig.9　Sample extraction results

4.3　文本分类模型构建

4.3.1　文本分类模型

对政策注意力分布进行分析的核心就是确定目前所有的政策措施的分布，对已发布的管理措施进行分类后即可清晰地看出目前政府在制定政策时注意力的重点分布，本研究采用基于 Transformer 的方法进行措施文本的分类。

基于 Transformer 的文本分类流程主要包括如下几

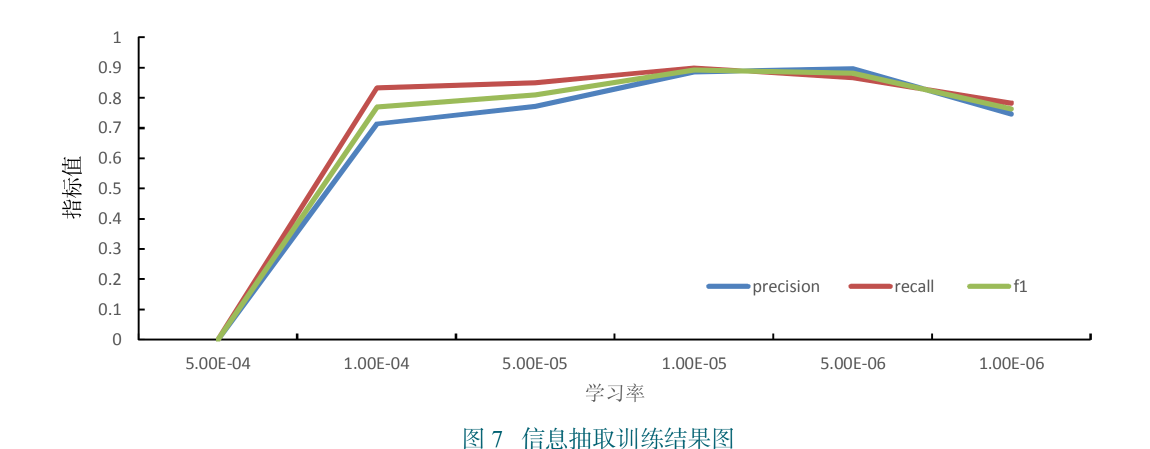


图 7　信息抽取训练结果图

Fig.7　Information extraction training effect

个部分：Embedding 层，这一层获取文本的向量表示；在 Positional Encoding 层添加文本的位置信息，包括文本的绝对位置和相对位置；将输入的 Embedding 乘以一个矩阵产生 Query、Key、Value，基于 Q、K、V 利用点积计算 Q 与 K 的相似度，再放入 Softmax 中输出当前 Q 与各个 K 的相似度；将 Q、K、V 在每个维度上切成 h 份，分别进行点乘缩放，并将最终结果合并输出[15]。

4.3.2　模型训练微调

（1）语料标注。本部分模型训练的目的为找到最优的措施分类模型，首先对语料文本进行标注，本次训练选定 1 000 条语料文本进行标注，按照 8:2 的比例划分为训练集和测试集，将语料输入进模型进行训练微调，标注样例如表 3 所示。

表 3　分类模型标注语料样例

Table 3 Samples using the classification model in the annotated corpus

文本	类别
围绕打造中国北方旱地西红柿之乡目标	品牌定位与规划
提升苗种繁育技术	品牌核心能力提升
创新宣传推广模式	品牌营销传播
举办“安化黄精”宣传标语口号征集活动	品牌营销传播
完善农业支持保护政策出台	品牌保护
推动实施保护工程项目产品占获证产品的 30%以上	品牌保护

（2）实验设置。对于措施文本多分类问题，本研究使用了机器学习中被广泛使用的深度学习框架 Pytorch，具体的训练框架及硬件版本如表 4 所示。

表 4　模型训练框架与硬件

Table 4 Model training framework and hardware

框架与硬件	版本
Python	3.9
Pytorch	1.11.0
GPU	RTX3080
CUDA	11.6

（3）训练调优。为了使训练的模型获得最佳的分类效果，本研究基于已构建语料选取不同的实验参数进行比较训练，这里同样选取对分类效果影响较大的学习率参数进行训练调整，在其他参数均为默认值的前提下，设置不同的学习率进行实验，在学习率为 5e-4 到 1e-5 的范围内进行训练，得到结果如图 10 所示，横坐标为学习率的值，纵坐标为不同学习率下各项实验指标的结果。

训练结果表明，当学习率在 1e-4 时，$F1$ 值达到最优。所以选取学习率为效果较优的 1e-4 进行后续措施分类。

5　以烟台苹果为例分析

利用已训练优化好的措施诊断模型参数对措施库中的文本进行措施诊断。将措施文本按条输入已构建的模型中，逐条输出措施分类结果，对分类结果进行统计得到措施类型分布。以烟台苹果为例，经过第一步措施抽取共获得关于烟台苹果措施文本 1 142 条，将其作为输入文本输入进分类模型，最终文本分类结果

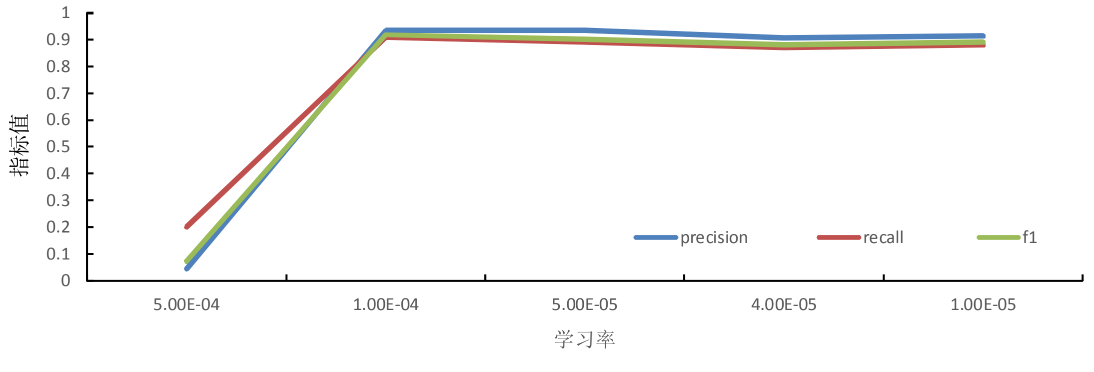


图 10　文本分类训练结果图

Fig.10　Text classification model training effect

DOI：10.13998/j.cnki.issn1002-1248.23-0312

示例如表 5 所示。

表 5 烟台苹果品牌管理措施分类结果

Table 5 Classification results of the brand management measures of Yantai apples

序号	文本	类别
1	形成品质差异化的品牌特征	品牌定位与规划
2	形成区域产业企业农户的合纵连横	品牌定位与规划
3	利用大数据云计算物联网等新一代信息技术	品牌核心能力提升
……	……	……
1139	讨论当前国际适用的地理标志保护模式	品牌保护
1140	扶持一笔资金构建一套服务体系的五个一机制	品牌管理
1141	构建近万人的营销网络和体系	品牌营销传播
1142	加强农产品归属管理	品牌管理

将所有的分类结果进行汇总，得到烟台苹果的政策注意力分布，可知烟台苹果在品牌管理过程中对于“品牌定位与规划”“品牌核心能力提升”比较重视，而应该加强“品牌营销传播”“品牌保护”“品牌管理”方面的措施发布，结果如表 6 所示。

表 6 烟台苹果品牌管理措施分类汇总

Table 6 Classification summary of the brand management measures of Yantai apples

序号	类别	数量/个
1	品牌定位与规划	469
2	品牌核心能力提升	362
3	品牌营销传播	75
4	品牌保护	83
5	品牌管理	153

5.1 烟台苹果政策注意力分布

根据分类结果绘制地理标志农产品品牌管理政策注意力分析图，如图 11 所示，横轴代表所抽取的措施属于该类措施的数量，纵轴代表措施分类，据图可发现目前烟台苹果政策注意力呈现出较集中的形势，各地政府的注意力集中在品牌定位与规划和品牌核心能力提升两部分，如提升农产品质量安全、升级产业链、加大金融融资、建立果园、提高农业应用水平、多产业结合等方式，《指南》指出品牌发展不能忽视品牌营销及保护，同时也不能实行模糊的品牌管理主体职能划分，不利于加强品牌监管。

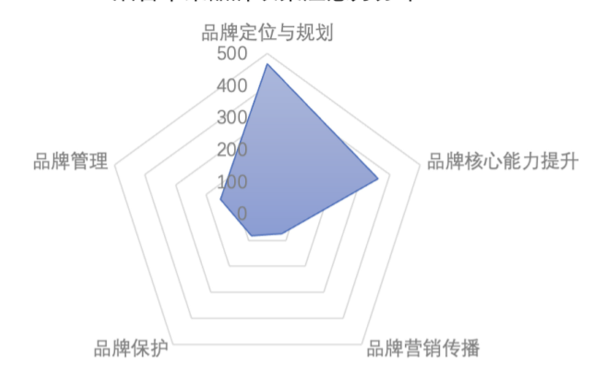


图 11 烟台苹果品牌政策注意力分布图

Fig.11 The attention distribution map of Yantai apple brand policies

6 结 论

本文基于文本挖掘技术并结合地理标志农产品品牌管理领域知识，构建了地理标志农产品政府管理注意力研究模型，涉及文本获取–处理–表示–抽取–分析全流程。首先我们设置了文本爬取策略以构建领域语料库，在命名实体识别部分采用基于领域词典的方式提高了专有名词的抽取效率，在措施提取部分我们采用 UIE 中文信息通用抽取框架，通过结合领域知识进行微调训练，最终 *F*1 达 0.89，并基于该模型构建了地理标志农产品品牌管理措施知识库。在文本分析阶段，我们采用基于 Transformer 的文本分类模型，最终完成研究模型的构建。

完成构建模型之后，我们选取烟台苹果品牌的文本进行的政策注意力分析，发现目前烟台苹果政府管理重点在品牌定位规划和核心能力提升部分，对品牌管理、保护营销涉足较少，为政府优化品牌发展政策提供数据支持，促进品牌发展。

参考文献：

[1] 农业农村部. 关于开展2022年农业品牌精品培育工作的通知: 农办市[2022]10号.[2022-06-30]. http://www.moa.gov.cn/govpublic/SCYJJXXS/202206/t20220630_6403818.htm.

[2] 詹君恒. 原产地效应及其作用机制: 一个文献综述[J]. 中国市场, 2019(16): 19-22.
ZHAN J H. Origin effect and its mechanism: A literature review[J]. China market, 2019(16): 19-22.

[3] 侯雅洁. 农业品牌建设将走进标准引领新时代[J]. 农业知识, 2022(10): 40, 42.
HOU Y J. Agricultural brand construction will enter the standard and lead the new era[J]. Agriculture knowlege, 2022(10): 40, 42.

[4] 房玲玲, 杨颖秀. 师范生公费教育政策文本分析及政策建议[J]. 延边大学学报(社会科学版), 2020, 53(1): 124-132, 144.
FANG L L, YANG Y X. Text analysis of the free education policy for normal students and policy suggestion[J]. Journal of Yanbian university (social sciences), 2020, 53(1): 124-132, 144.

[5] 宋娇娇, 徐芳, 孟溦. 中国科技评价政策的变迁与演化：特征、主题与合作网络[J]. 科研管理, 2021, 42(10): 11-19.
SONG J J, XU F, MENG W. Changes and evolutions of science and technology evaluation policies in China: Characteristics, subjects and collaborative networks[J]. Science research management, 2021, 42(10): 11-19.

[6] 王长征, 彭小兵, 彭洋. 地方政府大数据治理政策的注意力变迁——基于政策文本的扎根理论与社会网络分析[J]. 情报杂志, 2020, 39(12): 111-118.
WANG C Z, PENG X B, PENG Y. Attention evolution of big data governance policies of local government - Based on grounded theory and social network[J]. Journal of intelligence, 2020, 39(12): 111-118.

[7] 俞兆达. 新时代中国高等教育质量政策焦点的质化考察[J]. 当代教育论坛, 2022(2): 1-11.
YU Z D. A qualitative study on the focus of China's higher education quality policy in the new era[J]. Forum on contemporary education, 2022(2): 1-11.

[8] 徐宛笑, 李文欣, 黄栋. 中国—东盟文化产业合作中地方政府注意力分配研究——基于地方政府政策文本分析[J]. 东南亚纵横, 2022(5): 93-104.
XU W X, LI W X, HUANG D. A study of the attention allocation of local governments in China-ASEAN cultural industry cooperation: In the case of local government policy texts[J]. Crossroads: Southeast Asian studies, 2022(5): 93-104.

[9] 朱斌, 宁伟, 张静雅, 等. 我国基本医保门诊慢特病政策分析——基于内容分析法[J]. 中国卫生政策研究, 2022, 15(1): 43-49.
ZHU B, NING W, ZHANG J Y, et al. Policy analysis of basic health insurance chronic disease and special critical diseases security in China: Based on content analysis[J]. Chinese journal of health policy, 2022, 15(1): 43-49.

[10] SUN Y, WANG S H, FENG S K, et al. ERNIE 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation[J]. arXiv preprint arXiv:2107.02137, 2021.

[11] DAI Z H, YANG Z L, YANG Y M, et al. Transformer-XL: Attentive language models beyond a fixed-length context[J]. arXiv preprint arXiv:1901.02860, 2019.

[12] CHEN X R, YUILLE A L. Detecting and reading text in natural scenes [C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR. Piscataway, New Jersey: IEEE, 2004: II.

[13] MILWARD D, THOMAS J. From information retrieval to information extraction[C]// Proceedings of the ACL-2000 workshop on Recent advances in natural language processing and information retrieval held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics - Morristown, NJ, USA: Association for Computational Linguistics, 2000: 85-97.

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all You need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[15] 孙楚权. 基于Transformer的文本分类算法及应用[D]. 烟台: 山东工商学院, 2021.
SUN C Q. Algorithm and application of text classification based on transformer[D]. Yantai: Shandong Technology and Business University, 2021.

DOI：10.13998/j.cnki.issn1002-1248.23-0312

Construction and Application of the Attention Analysis Model of Brand Management Policies of Agricultural Products with Geographical Indications

HUO Mengjia[1], LIU Juan[1,2*], Huang Jie[1]
(1. Institute of Agricultural Information, Chinese Academy of Agricultural Sciences, Beijing 100080;
2. Key Laboratory of Agricultural Big Data, Ministry of Agriculture, Beijing 100080)

Abstract: [Purpose/Significance] Geographical indications (GIs) are an important tool for local governments in China to carry out brand building of agricultural products. Brand management is a continuous systematic project involving multiple subjects. Among them, the problem of government policy attention in the field of brand management of agricultural products with GIs deserves in-depth study. This paper aims to construct a policy attention analysis model in the field of brand management of GI agricultural products based on natural language processing technology. This model provides technical support for local governments to explore the status quo of local GI agricultural products' brand management, analyze the distribution of policy attention, and assist in optimizing strategies to promote their products' brand development. [Method/Process] The study is focused on the distribution of attention paths of GI agricultural products' brand management policies from the perspective of the local government: an analysis model of brand management policies of GI agricultural products was constructed in order to support the local government to carry out the analysis of the status quo of local agricultural products' brand management and policy optimization, and provide decision-making support for the optimization of brand management measures of the local agricultural products. First, this paper built a basic corpus based on the Python crawler technology, collected authoritative public information on the Internet, utilized the domain dictionary and UIE general information extraction framework to extract the text of management measures published in local government policies, and built a database of brand management measures of GI agricultural product. Second, a classification model of brand management measures of GI agricultural products based on the Transformer model was constructed. Third, this paper built a classification model based on the Transformer model, which can classify the extracted brand management measures of agricultural products and construct the policy attention distribution map. Finally, based on the policy attention distri-bution given by the model, we can find the brand management bottlenecks and recommend countermeasures to solve the bottlenecks. [Results/Conclusions] This paper takes Yantai apples as an example for model validation. After extracting and categorizing Yantai apples' brand management data, it is found that the policy attention of Yantai apples is more con-centrated, and the measures are highly similar, with 41.1% of the text of the measures focused on the part of brand positioning and planning, 31.7% on the part of brand competitiveness enhancement, and less than 10% on the part of brand marketing and protection. It can be seen that the brand effect of Yantai apples with GIs has not been well utilized.

Keywords: geographical indications of agricultural products; natural language processing; policy attention; information extraction; text classification